DOCUMENT RESUME

ED 264 272 TM 850 758

AUTHOR Embretson, Susan Whitely
TITLE Component Latent Trait Models for Test Design.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jul 82
NOTE 24p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Aptitude Tests; Difficulty Level; Estimation (Mathematics); Hypothesis Testing; *Latent Trait Theory; *Mathematical Models; *Maximum Likelihood Statistics; *Measurement Techniques; Performance; *Test Construction; Test Items
IDENTIFIERS *Item Parameters

ABSTRACT

Latent trait models are presented that can be used for test design in the context of a theory about the variables that underlie task performance. Examples of methods for decomposing and testing hypotheses about the theoretical variables in task performance are given. The methods can be used to determine the processing components that are involved in item performance. Three component latent trait models for underlying theoretical variables are described along with their maximum likelihood estimators. The item parameters can be used for item banking according to the influence of the underlying processing variables on item difficulty. Such estimators permit the test developer to choose items that represent specified information processing demands for the examinee. In this manner, what is measured by an aptitude test can be explicitly designed by specifying difficulty levels in the underlying processing components. The need for meta component latent trait models was also considered. It was shown that both items and persons vary on metacomponent parameters and that these parameters are important for the predictive validity of an aptitude test. The main conclusion to be drawn from these studies is that metacomponent latent trait models are needed to estimate more fully the processing abilities that underlie aptitude. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED264272

Component Latent Trait Models for Test Design

Susan Embretson (Whitely)
University of Kansas

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☑ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Research on aptitude tests has changed considerably in the last decade. The infusion of cognitive psychology into aptitude research has revitalized the field. Research on the cognitive components of aptitude (Carroll, 1976; Pellegrino & Glaser, 1979; R. Sternberg, 1977), as well as the cognitive correlates of aptitude (Hunt, Lunneborg, & Lewis, 1975), not only has changed the content of aptitude theory but also has influenced the type of data that is deemed relevant.

Cognitive psychology differs markedly from psychometrics on the role of the stimulus in task performance. Cognitive psychology experiments often employ within-subjects factorial designs in which stimuli are systematically manipulated to represent different levels of theoretical variables. Other theoretical variables that could influence performance are either held constant over the set or counterbalanced to eliminate bias. These experiments are like psychological tests in that many problems of a single task type are presented. However, the goal is to decompose the stimulus factors in the task that influence performance.

Cognitive component analysis of aptitude seeks to decompose the factors that influence performance on aptitude test items. A wide variety of the item types that appear on popular tests have been studied experimentally. For example, linear syllogisms (Sternberg & Weil, 1981), series completions (Butterfield, in press), and spatial problems (Pellegrino, Mumaw, & Cantony, in press), as well as many other item types, have been studied in recent research on cognitive components. The factors that have been identified on these tasks include the processes, strategies, and knowledge stores that underlie performance.

Cognitive component decomposition of aptitude offers a new approach to psychological measurement. This approach is test design, in which the qualities that are measured by a test are operationalized by the design of the test stimuli. That is, just like an experimenter who designs tasks to test hypotheses, an item writer manipulates the stimulus features of an item to represent specified theoretical constructs. Test design may be applied to many substantive areas and linked directly to psychometrics (see Embretson, in press-b).

The test design approach involves qualitatively different assumptions about the nature of construct validation research. Traditionally, the construct validity of a measure is assessed through the relationship of individual differ-

TM 850 758

ences on the test to other measures. Recently, Embretson (in press-a) has elaborated on two separate goals in construct validation research-construct representation and nomothetic span. Embretson (in press-a) hypothesized that the shift of psychological research to structuralism permits construct representation to be studied separately from nomothetic span. In Embretson's (in press-a) conceptualization of construct validity, construct representation is assessed from task decomposition data, while nomothetic span is assessed from individual differences data. That is, the theoretical constructs that are represented in performance may be studied independently from the utility of the test as a measure of individual differences. Thus, the construct validity of the test depends, in part, on the representation of the underlying constructs in the item task.

The goal of the current paper is to present three latent trait models that can be used for test design. Estimating the parameters for these models depends on applying a method for task decomposition. Thus, prior to presenting the latent trait models, two methods for task decomposition will be presented, along with examples that illustrate their relevance for test design. Then, the three latent trait models will be presented. These are (1) the linear logistic latent trait model (Fischer, 1973); (2) the multicomponent latent trait model (Whitely, 1980d); and (3) the general component latent trait model (Whitely, 1980a). The latter is a generalization that includes the other two models. Last, the need for more complex latent trait models to fully assess the important cognitive components of aptitude will be examined. That is, the potential contribution of metacomponent latent trait models to test validity will be explored.

Methods for Task Decomposition

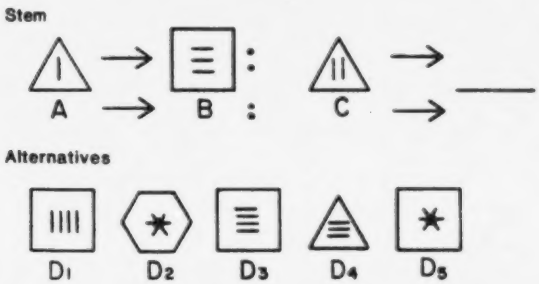
Methods for task decomposition are a major tool in contemporary research on cognitive components. The methods that are applied to decompose tasks may also be applied to the design for test stimuli. Two popular methods for task decomposition are (1) the method of complexity factors and (2) the method of subtasks. An example of how task decomposition methods can be used for test design will be presented for each method.

In the method of complexity factors, each item is manipulated and/or scored on one or more factors that represent the item's position on underlying theoretical variables. This method has been applied to attitude and personality items (Cliff, 1977; Cliff, Bradley, & Girard, 1973), as well as to a wide variety of cognitive tasks, such as linear syllogisms, geometric analogies, series completion problems, and spatial rotation items.

Figure 1 presents an example of a geometric analogy (Whitely & Schneider, 1981) that represents the method of complexity factors. Two processing events have been indicated as having major influence on task difficulty (Mulholland, Pellegrino, & Glaser, 1980; Whitely & Schneider, 1981). These are (1) encoding complexity, which depends on the number of elements in the A term in the analogy and (2) transformational complexity, which depends on the number of transformations that are required to convert A to B. In Figure 1 the A term contains two elements (the triangle and the line) and the A to B conversion requires three transformations (a shape change of the external element, an increase in the num-

ber of internal elements and a 90° rotation of the internal elements). Whitely and Schneider (1981) found that two different types of transformation had opposing influence on item difficulty. Distortions (change in shape or number) were positively related to accuracy, while displacements (rotations) were negatively related to accuracy.

Figure 1
A Geometric Analogy, Similar to an Item
on the Cognitive Abilities Test



These findings indicate that the test developer can control item difficulty by systematically varying the number of elements and the number and type of transformations in the item stimuli. An easy item would have one or two elements and a distortion transformation. A difficult item would have several elements and one or more displacement transformations. Thus, the test developer can fashion items to achieve desired levels of difficulty.

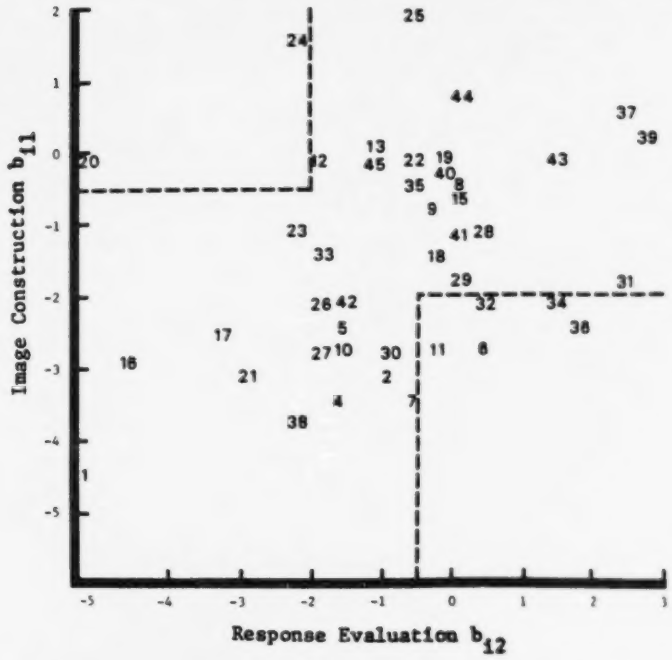
In contrast to the method of complexity factors, the method of subtask responses requires the theoretical variables to be identified from a series of subtasks that have been constructed from the items. Table 1 presents a verbal analogy item that is similar to items on the verbal section of the Cognitive Abilities Test. The total item, as presented on the test, is given at the top. Two components that have been supported by previous experimental research on verbal analogies are Rule Construction and Response Evaluation (Pellegrino & Glaser, 1979; Whitely, 1980c; Whitely & Barnes, 1979). These are represented by the two subtasks in Table 1. Notice that although Response Evaluation is sequentially dependent on Rule Construction, supplying the rule in the subtask makes possible independent assessment of these components. Thus, for each item, examinees respond to the total item as well as to the subtasks that represent processing components.

By using the psychometric models to be described below, item difficulty on the components underlying the subtasks can be calibrated on a common scale. Figure 2 presents a scatterplot of the item parameters on the two components. It can be seen that item difficulty on the two components is not highly related. Thus, it is possible to design tests that reflect predominantly the influence of one component or the other. For example, items that are easy on Response Evaluation but difficult on Image Construction would measure abilities on the latter. The test developer could select the items in the lower right corner to meet this specification.

Table 1
Subtask Set for Verbal Analogy Components

Total Item
Cat : Tiger :: Dog : _____
(a) Lion (b) Wolf (c) Bark (d) Puppy (e) Horse
Rule Construction
Cat : Tiger :: Dog : _____
Rule ? _____
Response Evaluation
Cat : Tiger :: Dog : _____
(a) Lion (b) Wolf (c) Bark (d) Puppy (e) Horse
Rule: A large or wild canine

Figure 2
Scattergram of Image Construction Difficulty by Response Evaluation Difficulty for 45 Verbal Classification Items



Desiderata for Psychometrics for Test Design

The indices that are derived from classical test theory or latent trait theory do not reflect the stimulus properties of a test item with respect to specified factors. There are several desiderata for test theory models that can be applied to test design. First, the method must be capable of testing hypotheses about the specification factors. Obviously, a viable specification system is one that is highly related to item difficulty. However, hypothesis testing

about the factors in items is also crucial to establishing a theory of the item task. An item specification system is implicitly a theory of the task so that it should be evaluated by the hypothesis-testing methods that are applied to other theories. Second, the model must have parameters to describe the difficulty of the items on the underlying factors. The unidimensional latent trait models that are popular in test development do not have this property, since the items are calibrated for only one dimension--the largest common factor in the items. A model that allows designation of the difficulty factors according to an a priori specification is required. Third, measurements of persons must be included in the model. The need for person measurements is self-evident, since the goal of aptitude testing is to measure individual differences. Fourth, the model should specify the relationship between the item parameters and the person abilities. Optimally, the test design approach involves selecting from a calibrated item bank for a certain measurement goal. It is essential that the influence of item parameters on person abilities is well specified in the model.

Component Latent Trait Models

This section presents three component latent trait models that can be used to test hypotheses about construct representation and to assess factors for test design. These are (1) the linear logistic latent trait model, (2) the multicomponent latent trait model, and (3) the general component latent trait model. The latter, a generalization of the other two, can handle more complex data about cognitive processes.

The Linear Logistic Latent Trait Model

The model. The linear logistic latent trait model (LLTM) is a unidimensional model in which components are identified from item scores on complexity factors that are postulated to determine item difficulty. To understand how components are identified, consider the geometric analogy presented in Figure 1, which is similar to items on the nonverbal section of the Cognitive Abilities Test. A recent study (Whitely & Schneider, 1981) compared three cognitive models of geometric analogies, using the LLTM. All three models specify complexity factors in processing the item that influence response difficulty.

The scores of the items on the complexity factors identify the components in an LLTM. The model can be examined by considering three equations. The first equation is the mathematical model for task processes. Here, a linear model of the complexity factors, c_{im}, multiplied by their difficulty, η_m, predicts item difficulty, b_i^*.

$$b_i^* = \sum_m c_{im}\eta_m + d \ , \qquad [1]$$

where

c_{im} = the complexity of factor m in item i;

η_m = the difficulty of complexity factor m; and

d = a normalization constant.

The second equation presents the latent trait model for individual differences,

which is the Rasch latent trait model,

$$P(x_{ij}=1|\theta_j,b_i) = \frac{e^{(\theta_j-b_i)}}{1+e^{(\theta_j-b_i)}} \qquad [2]$$

where θ_j = ability for person j and b_i = difficulty for item i.

Equation 3 combines these two models to give the LLTM as follows:

$$P(x_{ij}=1|\theta_j,\eta_m,d) = \frac{e^{(\theta_j-(\sum_m c_{im}\eta_m + d))}}{1+e^{(\theta_j-(\sum_m c_{im}\eta_m + d))}} \qquad [3]$$

If the number of complexity factors equals the number of items, and each item contains only one complexity factor, then the LLTM is equivalent to the Rasch latent trait model. When the number of η_m is less than the number of items, the LLTM is a linearly constrained model of item difficulty.

A major advantage of the LLTM is the possibility of comparing alternative models of item difficulty by χ^2 difference tests based on the log likelihood of the data, given the model. For example, the fit of any restricted model of the task components can be compared to the fit of the Rasch model, which can be regarded as a saturated model of item difficulty. Further, if alternative models of task components are hierarchically nested, then direct comparisons between the models are also possible. Thus, hypothesis testing to establish a valid model of the task complexity factors is an important capability of the LLTM.

Another important aspect of the model is that of parameters describing each item by component complexity rather than just item difficulty. These parameters can be useful in item banking, so that the contribution of a processing complexity factor to each item is systemically specified. Notice, however, that the model is unidimensional, since only one ability parameter is specified for each person.

Estimation. Fischer (1973) derived conditional maximum likelihood estimators for the item parameters of the LLTM, η_m. Although conditional maximum likelihood estimators are statistically superior to unconditional estimators for several reasons (Fischer, 1981), they are impractical for large sets ($I > 60$). Thus, unconditional maximum likelihood estimators are useful for LLTM item parameters.

The first derivative of the log likelihood function for unconditional maximum likelihood estimation is

$$\frac{\partial L}{\partial b_i} = \sum_j \left[x_{ij} - \left(1 + e^{-(\theta_j - b_i)}\right)^{-1} \right]. \qquad [4]$$

Thissen (1981) has shown that the first derivative of the log likelihood function L for the LLTM with respect to η_m may be obtained from

$$\frac{\partial L}{\partial \eta_m} = \sum_i c_{im} \left[\frac{\partial L}{\partial b_i}\right] . \qquad [5]$$

Combining Equation 4 with Equation 5 gives the first derivative of the unconditional log likelihood function with respect to η_m

$$\frac{\partial L}{\partial \eta_m} = \sum_i c_{im} \sum_j \left[x_{ij} - \left(1 + e^{-(\theta_j - b_i^*)}\right)^{-1}\right] , \qquad [6]$$

where b_i^* is defined as in Equation 1.

Multicomponent Latent Trait Model

The model. The multicomponent latent trait model (MLTM) is a multidimensional model in which components are identified from subtasks that represent the processing components in item solving. Like many information processing models for complex tasks (e.g., Hunt, 1976), it is assumed that information from several component events is required to solve the item. The relationship between the component events may be either (1) independent, where the processing or outcome of one event does not influence any other event, or (2) sequentially dependent, where information from a component event provides prerequisite information for processing on later events.

A MLTM uses subtask data to identify the components. The mathematical model of processes in the MLTM links the component responses to the total item. Equation 7 presents a mathematical model for independent components in which the response probability for the total item is the product of the component likelihood:

$$P(x_{ijT}=1) = a \prod_k P(x_{ijk}=1) + g \left[1 - \prod_k P(x_{ijk}=1)\right] . \qquad [7]$$

where

$P(x_{ijT}=1)$ = the probability that the composite task is correct for person j on item i,

$P(x_{ijk}=1)$ = the probability that the subtask for component k is correct for person j on item i,

a = the probability that an item is solved when the component information is available, and

g = the probability of solving an item when the component information is not available.

Unlike the original MLTM, the model includes parameters for application of the component information, a, which represents metacomponent or executive functioning, and for an alternative method for solving the item, g, such as guessing or rote association to the stem. Other mathematical models are possible (e.g., Whitely, 1980d), but all models relate the component likelihoods to the full item likelihood.

As for the LLTM, the latent trait model for individual differences is the 1-parameter logistic latent trait model, as presented in Equation 2. However, in the MLTM the latent trait models are given for component subtask responses rather than for the total item. The MLTM specifies that responses to the subtasks depend on the ability of person j on component k and the difficulty of item i on component k, as follows:

$$P(x_{ijk}=1|\theta_{jk},b_{ik}) = \frac{e^{(\theta_{jk}-b_{ik})}}{1+e^{(\theta_{jk}-b_{ik})}} , \qquad [8]$$

where θ_{jk} = the ability of person j on component k and b_{ik} = the difficulty of item i on component k. The LLTM, in contrast, is a latent trait model for responses to the total item and does not model component responses.

The full model, presented in Equation 9, combines the latent trait model with the mathematical model. It can be seen that the total item response is conditional on K component abilities as well as on K component item difficulties.

$$P(x_{ijT}=1|\underset{\sim}{\theta}_j,b_i) = (a-g)\prod_k \frac{e^{(\theta_{jk}-b_{ik})}}{1+e^{(\theta_{jk}-b_{ik})}} + g . \qquad [9]$$

where $\underset{\sim}{\theta}_j$ = the vector of k component abilities for person j and b_i = the vector of k and component difficulties for item i.

Although typical test data (with the notable exception of linked items with a common stem) only occasionally assess subtask responses, there are several reasons why it may be useful to obtain such data as part of test development. First, the various processing components from which information is required for item solution are theoretically distinct. Experimental cognitive research has supported the independence of components within a task by additive factor (S. Sternberg, 1969) or subtractive factor (Pachella, 1974) modeling methods. Second, if the components are sufficiently elementary, they should generalize across tasks and possibly account for differential patterns of correlations in performance on separate types of items (Carroll, 1974). Third, individual differences on different components correlate only moderately and show differential validity in predicting performance on other tasks (R. Sternberg, 1977; Whitely, 1981). Fourth, component difficulties are sometimes not highly correlated in item sets, so that it is possible to select items of the same type that measure different component abilities. Consider for example, a two-component item, such as presented in Table 1. If items are so easy on one component that nearly everyone has a high probability of executing it correctly, then it can be shown that the likelihood of correctly answering the items is well described by the regression of the response likelihoods on the other component ability (Whitely, 1981).

The relationship of the MLTM parameters to the joint response to the components, C_i, and the total item T, is explicated more completely by considering the probability sample space. There are 2^{K+1} possible response patterns. Table 2 shows the eight response patterns for a two-component item, along with an expression for the probability of the pattern from the MLTM. It can be seen that the a and g parameters link the component response to the total item, while the other symbols represent the probability of the component response patterns, which vary systematically over persons and items, according to the 1-parameter logistic latent trait model.

Estimation. No estimators were developed in Whitely (1980d) for the MLTM. However, given a probability space such as specified in Table 2, the likelihood of any response pattern is given by

$$P(x_k, x_T) = \left[g^{x_T}(1-g)^{1-x_T} \right]^{1-\Pi x_k} \left[a^{x_T}(1-a)^{1-x_T} \right]^{\Pi x_k} \left[\Pi_k P_{x_k}^{x_k} Q_{x_k}^{1-x_k} \right] , \qquad [10]$$

where

x_k = vector of responses of person j to components for item i,

x_T = response of person j to total item i, and

$P_{x_k} = P(x_{ijk}=1 | \theta_{jk}, b_{ik})$.

Notice that the entry of the parameters a and g into the likelihood depends on the value of $\pi_k x_k$ and that $\pi_k x_k$ equals 1.0 only if all component outcomes are correct. Note also that a contributes to the log likelihood only if all the components are executed correctly, while g contributes when at least one component is incorrect. This pattern is specified in Table 2.

The likelihood of the data set can be obtained by multiplying the response likelihoods over persons and items. Since neither a nor g vary over persons or items, it can be concluded immediately from well-known theorems on the binomial distribution that their maximum likelihood estimators are the relative frequencies

$$a = \frac{\sum_{ji} (\Pi x)_{k\,k}\, x_T}{\sum_{ji} \Pi x_{k\,k}} \qquad [11]$$

and

$$g = \frac{\sum_{ji} (1 - \Pi x)_{k\,k}\, x_T}{\sum_{ji} (1 - \Pi x)_{k\,k}} . \qquad [12]$$

Thus, a is given by the relative frequency of correctly answering the item when all components are executed correctly, while g is given by the relative frequency of correctly answering the total item when at least one component is executed incorrectly.

Table 2
Frequencies and Conditional Probabilities for
Joint Response Patterns on Verbal Analogies

C_1	C_2	T	f	$P(x_T = 1 \mid x_k)$	Notation
1	1	1	1864	.84	$a\ P_{x_1} P_{x_2}$
1	1	0	351	.16	$(1-a)\ P_{x_1} P_{x_2}$
0	1	1	518	.50	$g\ Q_{x_1} P_{x_2}$
0	1	0	518	.50	$(1-g)\ Q_{x_1} P_{x_2}$
1	0	1	84	.45	$g\ P_{x_1} Q_{x_2}$
1	0	0	101	.45	$(1-g)\ P_{x_1} Q_{x_2}$
0	0	1	87	.28	$g\ Q_{x_1} Q_{x_2}$
0	0	0	221	.72	$(1-g)\ Q_{x_1} Q_{x_2}$

$$P_{x_k} \equiv P(x_{ijk}=1 \mid \theta_{jk}, b_{ik})$$

$$P_{x_k} = \frac{e^{(\theta_{jk}-b_{ik})}}{1 + e^{(\theta_{jk}-b_{ik})}}$$

$$P(x_T=1) = a \prod_k P_{x_k} + g\left[1 - \prod_k P_{x_k}\right]$$

$$P(x_k, x_T) = \left[g^{x_T}(1-g)^{1-x_T}\right]^{1-\prod x_k}\left[a^{x_T}(1-a)^{1-x_T}\right]^{\prod x_k}\left[\prod_k P_{x_k}^{x_k} Q_{x_k}^{1-x_k}\right]$$

The required derivative of the log likelihood for unconditional maximum likelihood estimation of the item parameters $\underline{b}$ is

$$\frac{\partial L}{\partial b_{ik}} = \sum_j \left[x_{ijk} - \left(1 + e^{-(\theta_{jk} - b_{ik})} \right)^{-1} \right] . \qquad [13]$$

Setting the derivative to zero leads to the well-known equations for unconditional maximum likelihood estimation of the 1-parameter logistic latent trait model (cf. Lord & Novick, 1968). As for other exponential families of distributions, estimation equations for the latent trait model can be obtained by equating the observed sufficient statistics with their expectancies, given the parameters (Andersen, 1980). In the current development, however, estimation requires I equations for <u>each</u> of K components of the MLTM. Notice that the item parameters for each component b_{ik} involve only the responses to the relevant subtask data, x_{ijk}. It can be seen that unconditional maximum likelihood estimators may be obtained independently from each subtask to maximize the log likelihood of the joint response pattern x_k, x_T.

A General Multifactor Latent Trait Model

<u>The model</u>. The preceding developments have shown that the LLTM and the MLTM differ substantially in component identification. LLTMs estimate difficulty of complexity factors that are related to item difficulty, while MLTMs estimate item and person parameters for component outcomes. The different methods of component identification make possible a meaningful unification of these two models.

Consider the verbal analogy that is presented in Table 1. This analogy was presented previously with the MLTM. Although the Response Evaluation component is identical to the previous example, the Rule Construction component is postulated to be influenced by the several processing complexity factors, c_{ikm}, that are listed in Table 3. These factors concern the difficulty of inferring the target relationship (i.e., Fist: Clench). The factors c_{i11} and c_{i12} are the ease of inferring the target relationship in the initial encoding of the relational pair and in the context of the unmatched term "Teeth," respectively. Previous research on analogies (R. Sternberg, 1977) as well as research in memory organization (Reitman, 1965) suggest that relational span is also positively related to item solving, since extraneous relationships can interfere with solving the analogy. In the current example, the factors c_{i11} and c_{i12} are measured by the mean number of relationships that are educed between the word pair when presented alone and in the context of the unmatched term, respectively. The factors c_{i14} to c_{i17} represent the relative frequency of various types of context effects in inferring the target relationship (i.e., selecting or combining initial relationships, inferring new relationships, and so forth).

Scores for each item on the complexity factors were obtained from other research studies on analogies (Embretson & Curtright, 1981). However, it is

Table 3
Complexity Factors in Analogical Reasoning Components

Rule Construction Component
Fist:Clench::Teeth: _____
Rule? _____

c_{ikm} = the complexity of factor m for component k on item i
c_{i11} = inference elicitation, the probability that the target relationship is educed from initial word pair
c_{i12} = relational network span, the number of relationships educed
c_{i13} = inference contextualization, the probability that the target relationship is educed in context of all three stem stimuli
$c_{i1T} - c_{i17}$ = type of contextualization effect

Response Evaluation Component
Fist:Clench::Teeth: _____
(1) Pull (2) Brush (3) Grit (4) Gnaw (5) Jaw
Rule: Angry reaction done with "teeth."

important to note that in this example the complexity factors have effects on the component information outcomes rather than on the total item response.

Equation 14 presents a model that specifies both processing complexity factors and processing component outcomes.

$$P(x_{ijT}=1|\underset{\sim}{\theta}_j,\underset{\sim}{\eta}_m,\underset{\sim}{d}) = \prod_k \left[\frac{e^{(\theta_{jk}-(\sum_m c_{imk}\eta_{mk} + d_k))}}{1 + e^{(\theta_{jk}-(\sum_m c_{imk}\eta_{mk} + d_k))}} \right] \qquad [14]$$

As for the MLTM the probability of the correct response to the intact item is conditional on a vector of component abilities, $\underset{\sim}{\theta}_j$, and component item difficulties. However, item difficulty for each component η_m is determined by a linear model of the complexity factors for the component c_{ikm}.

Equation 14 is a general multifactor latent trait model (GLTM) for response processes. If only one information outcome is measured for the item (i.e., the response to the intact test item), then the model is identical to the LLTM. In this case, complexity factors would be scored for the total item. The parameters a and g drop out of the model, since the response to the total item would be given by response to the single component outcome that is observed. If no complexity factors postulated for each component outcome, but several component outcomes are observed, then the model is MLTM. In this case, the Rasch model is specified for each component. However, for tasks with multiple information outcomes and processing complexity factors that influence these outcomes, the full model can be utilized.

Estimation. The likelihood of the joint response pattern, x_k, x_T, given the parameters of the model, is given by Equation 10, except that the component likelihoods are given by the LLTM for the component as follows:

$$P_{x_k} \equiv P(x_{ijk}=1|\theta_{jk},\eta_{mk},d_k) \quad . \qquad [15]$$

The estimators of $\underline{a}$ and $\underline{g}$ for the GLTM are the same as for MLTM, as given in Equation 11 and Equation 12. Since item difficulty is linearly constrained within a component for the LLTM, the first derivative of the log likelihood function with respect to η_{mk} is required for unconditonal maximum likelihood estimation of the items. Using the development given above for LLTM, it can be seen that the first symbolic partial derivative with respect to η_{mk} in GLTM is

$$\frac{\partial L}{\partial \eta_{mk}} = \sum_i c_{imk} \sum_j \left[x_{ijk} - \left(1 + e^{-(\theta_j - b_i^*)}\right)^{-1} \right] . \qquad [16]$$

A Fortran program, MULTICOMP (Whitely & Nieh, 1981) is available to estimate the parameters of the GLTM.

Future Directions: Metacomponent Latent Trait Models?

The component models that were presented above do not fully reflect the complexity of the information processes that are involved in task performance. Metacomponent variables that determine when to execute component processes and which processes to execute have great impact on problem solving. For example, problem-solving strategies are an important concept in problem-solving theory (Davis, 1973; Newell & Simon, 1978). Similarly, problem-solving strategies have long been thought to be major aspects of individual differences in intelligence, particularly for those theories of intelligence that emphasize adaptability (Pintner, 1921; R. Sternberg, 1979; Woodrow, 1921). Thus, on theoretical grounds, a complete model of information processing on intelligence test items should include strategy variables.

The MLTM that is presented in Table 2 postulates that individuals have equal likelihoods of applying the various strategies. That is, the strategy application parameter $\underline{a}$, and the parameter for successful application of other strategies, $\underline{g}$, do not vary over persons or items. As suggested above, this assumption is unwarranted on both theoretical and empirical grounds, since metacomponents are known to influence task performance. Thus, to fully represent processing, the strategy application parameters need to vary over persons or items.

A metacomponent latent trait model would include strategy application parameters for persons or items. However, estimation of these parameters will be complex. Returning to Table 2, it should be obvious that the symbolic partial derivative with respect to $\underline{a}$, for example, will not be simple if variability for either persons or items is included in the model. That is, the estimation of such parameters will depend on the outcome of the other parameters, θ_j and b_i.

The parameter a can only be estimated from response patterns in which both components are correct, as in the first two response patterns in Table 2. Not only will the estimation algorithm necessarily be complex, but also estimation error will vary as a function of the other component responses. For example, for persons with few accurate component outcomes, the a parameter will not be estimated reliably, since little information about the parameter will be available.

An obvious question at this point is the potential utility of developing the estimators for the more complex metacomponent models. Two questions about metacomponent parameters need to be addressed: (1) Do items and persons vary in propensity for applying the various components? and (2) Do individual differences in metacomponents contribute to the criterion-related validity of an aptitude test? To answer these questions, data from two studies on verbal aptitude (Whitely, 1980, 1982) were reanalyzed to include the metacomponent parameters.

A reanalysis of data originally collected by Whitely (1980) shows the variability among items and persons in two metacomponents that could be estimated with the latent trait model in Table 2. These are application of the rule-oriented strategy, a, and application of other strategies, such as guessing, g. Figure 3 shows frequency distributions of the a parameters for two item types, verbal analogies, and verbal classifications. In this analysis, a was computed as the conditional probability that examinees would solve the total item when the component information was available. Admittedly, this estimator of a is crude, but it does provide at least some indication of its nature. It can be seen in Figure 3a that the parameter values tend to be high on both item types but that examinees vary widely on the parameter. It is not clear to what extent this distribution reflects differing degrees of accuracy of estimating a for individuals.

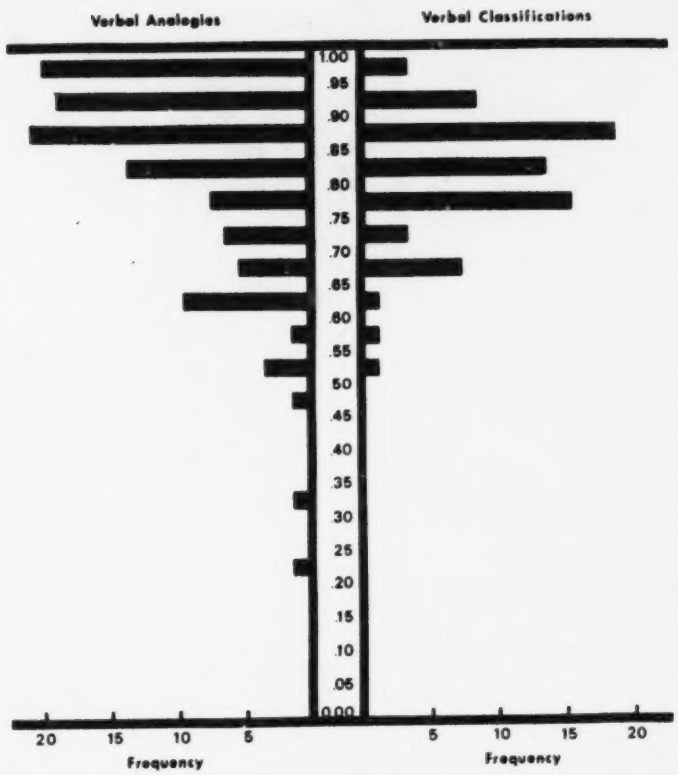
Figure 3b shows the distribution of the g parameter for examinees, computed as the conditional probability of solving the total item, given that the component information is not available. It can be seen that this value centers around .50 for both item types and that individuals vary widely in these values. Thus, for both strategy application and guessing, some individual differences are indicated. Figure 4 and Figure 5 present stem and leaf distributions of a and g parameters, respectively, for items. As for individuals, considerable variability is indicated.

A second study (Whitely, 1982) contains data on the contribution of metacomponent variables to test validity. The Whitely study examines the relationship of individual differences in strategy application to a major criterion for aptitude test validity, educational achievement. In this study, data on the achievement of 99 parochial high school students were collected, in addition to their performance on an analogical reasoning test and on several subtasks that represented components and metacomponents in solving analogies.

The contribution of strategy application parameters (i.e., a) and other strategies (g) were examined in separate analyses. In the Whitely (1982) analyses, individual differences in strategy application were examined for two strategies that led to analogy solving. There were (1) a rule-oriented strategy and (2) a response elimination strategy. The contribution of the strategy applica-

Figure 3
Frequency Distribution of Application (a) and Guessing (g)
Probabilities for Examinees on Two Item Types

(a) Application



(b) Guessing

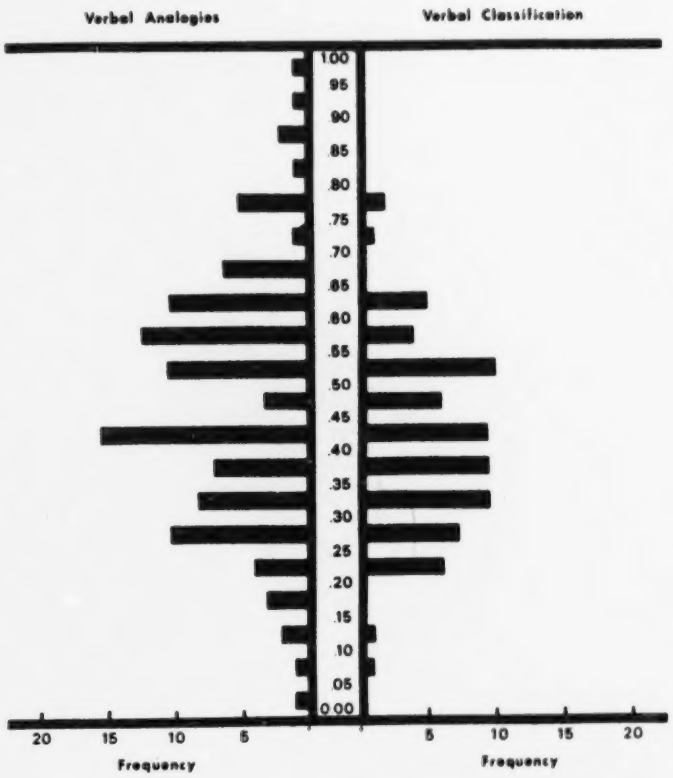


Figure 4
Stem and Leaf Distribution of Application (a)
Probabilities for Two Item Types

Verbal Analogies		Verbal Classifications
	1.000	006 028 039
030 029	.950	003 030
028 024 021 019 011 010 001	.900	002 011 016 021 023 042 043
033 032 025 022 018 017 014 009 002	.850	005 017 018 024 027 038
026 020 015 013 007 005 004	.800	004 009 013 015 019 026
034 016 003	.750	029 031 034 036 040
036	.700	008 035
027 023	.650	012 032 044
	.600	014 033 045
006	.550	020 022
037 031 012	.500	025
	.450	010
	.400	041
	.350	
	.300	
	.250	037
	.200	
	.150	
	.100	
	.050	
	0.000	

tion parameter to test validity was examined by structural equation models (Jöreskog, 1974).

In these models, individual differences in both applying and performing the components of the strategies were measured as independent variables. The dependent variables included performance on the analogical reasoning test as well as scores on eight area achievement tests. For both the rule-oriented strategy and the response elimination strategy, it was found that adding strategy application to the strategy performance variables significantly increased prediction of both analogical reasoning and achievement. The differences that were obtained by adding the strategy application variables to the covariance models were highly significant for both the rule-oriented ($\chi^2 = 65.57$, $p < .01$) and the response elimination strategy ($\chi^2 = 65.57$, $p < .01$). For both strategies the application

Figure 5
Stem and Leaf Distribution of Guessing (g)
Probabilities for Two Item Types

Verbal Analogies		Verbal Classifications
	1.000	
	.950	
	.900	
	.850	021
030	.800	003 006 011
025 021 011	.750	004
029 028 022 017	.700	002 042
018 009 002	.650	008 012 013 030 031
034 010 007 003	.600	028 036
001	.550	005
024	.500	017 029 034 038
015	.450	
033 012	.400	010 023 026
036 020 019 016 014 005	.350	009 018 019 035 043
031 023 013 004	.300	015 022 039
	.250	014 027 040
	.200	016
032 026	.150	045
006	.100	037
037 027	.050	020 024 044
	0.000	025 032 033 041

variable was significantly related to analogical reasoning (t = 7.58 and 2.07, respectively, for the rule-oriented and response elimination strategies), showing that strategy application is an important metacomponent for individual differences in analogical reasoning.

Table 4 presents data on the contribution of the application parameters to the prediction of achievement in several areas. Indices that are comparable to multiple regression analyses were obtained from the structural equation analyses. For each of the strategies and for the two strategies combined with a guessing strategy, Table 4 shows the F value for the application metacomponent and its incremental contribution to explaining variance of each achievement test, as well as the proportion of variance explained. The application metacomponent for the rule-oriented strategy significantly contributed to the valid-

ity for predicting Mathematics and Sources. The application metacomponent for the response elimination strategy significantly increased prediction for several achievement areas, including Reading Comprehension, Vocabulary, Language Use, Spelling, Social Science, Science, and Sources.

Table 4
Contribution of Metacomponent and Strategy Parameters to Predicting Achievement

Strategy and Achievement Area	Specification Accuracy Multiple R	Contribution of Metacomponent Reduction of F_β Error	(ΔR^2)
Rule-Oriented Strategy			
Reading Comprehension	.67	.28	.01
Vocabulary	.50	.02	.01
Language Use	.67	2.01	.02
Spelling	.38	.05	.00
Mathematics	.52	4.84*	.06
Social Science	.51	1.48	.02
Science	.74	.01	.00
Source	.66	9.18**	.09
Response Elimination Strategy			
Reading Comprehension	.71	5.42*	.05
Vocabulary	.66	14.21**	.14
Language Use	.73	6.81*	.06
Spelling	.52	8.35**	.11
Mathematics	.50	.55	.01
Social Science	.57	8.58**	.10
Science	.77	5.71*	.04
Source	.71	17.64**	.15
Guessing Strategy			
Reading Comprehension	.60	9.93**	.12
Vocabulary	.63	8.44**	.09
Language Use	.58	5.50*	.07
Spelling	.48	.05	.00
Mathematics	.69	2.61	.03
Social Science	.60	6.80*	.08
Science	.71	15.20**	.14
Source	.68	10.02**	.10

*p <.05.
**p <.01.

Table 4 also shows that adding the g parameter to the rule-oriented strategy and the response elimination strategy significantly increased the prediction of achievement in several areas. Thus, these data support the potential of individual differences in metacomponent variables as an important aspect of test validity. The metacomponent variables increased the prediction of achievement over the simple component performance variables.

The main conclusion to be drawn from these studies is that metacomponent latent trait models are needed to estimate more fully the processing abilities that underlie aptitude. Although the estimation of the metacomponent parameters will be complex, even the crude estimators that were used in the studies described above show clear contributions to aptitude test validity.

Conclusions

This paper has presented latent trait models that can be used for test design in the context of a theory about the variables that underlie task performance. Examples of methods for decomposing and testing hypotheses about the theoretical variables in task performance were given. The methods can be used to determine the processing components that are involved in item performance.

Three component latent trait models for underlying theoretical variables were described along with their maximum likelihood estimators. The item parameters can be used for item banking, according to the influence of the underlying processing variables on item difficulty. Such estimators permit the test developer to choose items that represent specified information processing demands for the examinee. That is, the test developer can select items that are difficult on some processes, but easy on others. In this manner, what is measured by an aptitude test can be explicitly designed by specifying difficulty levels in the underlying processing components.

The need for metacomponent latent trait models was also considered. It was shown that both items and persons vary on metacomponent parameters and that these parameters are important for the predictive validity of an aptitude test. Thus, metacomponent latent trait models should provide a better estimate of the abilities that are involved in test performance.

REFERENCES

Andersen, E. B. Discrete statistical models with social science applications. Amsterdam: North-Holland Publishing Company, 1980.

Barnes, G. M., & Whitely, S. E. Problem structuring processes for ill-structured verbal analogies. Memory and Cognition, 1981, 4, 411-421.

Butterfield, E. A model that predicts series completion accuracy. In S. Embretson (Whitely), (Ed.), Test Design: Contributions from psychology, education, and psychometrics. New York: Academic Press, in press.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." In L.B. Resnick (Ed.), The nature of intelligence. Hillsdale NJ: Erlbaum, 1976.

Cliff, N. Further study of cognitive processing models for inventory response. Applied Psychological Measurement, 1977, 1, 41-49.

Cliff, N., Bradley, P., & Girard, R. The investigation of cognitive models for inventory response. Multivariate Behavioral Research, 1973, 8, 407-425.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Davis, G. A. Psychology of problem solving: Theory and practice. New York: Basic Books, 1973.

Embretson, S. E. Test design: Contributions from psychology, education, and psychometrics. New York: Academic Press, in press. (b)

Embretson, S. E., & Curtright, C. Problem structure and response format on verbal analogies (Technical Report No. NIE 81-5). Lawrence: University of Kansas, Department of Psychology, December 1981.

Fischer, G. H. The linear logistic model as an instrument in educational research. Acta Psychologia, 1973, 37, 359-374.

Fischer, G. H. On the existence and uniqueness of maximum likelihood estimates in the Rasch model. Psychometrika, 1981, 46, 59-77.

Hunt, E. Mechanics of verbal ability. Psychological Review, 1978, 85, 109-130.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

J$reskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology. San Francisco: Freeman, 1974.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading MA: Addison-Wesley, 1968.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. Components of geometric analogy solution. Cognitive Psychology, 1980, 12, 252-284.

Newell, A., & Simon, H. A. Human problem solving. Englewood Cliffs NJ: Prentice-Hall, 1972.

Pellegrino, J. W., & Glaser, R. Cognitive correlates and components in the analysis of individual differences. Intelligence, 1979, 3, 187-214.

Pellegrino, J. W., Mumaw, R. J., & Cantoni, V. J. Analyses of spatial aptitude and expertise. In S. Embretson (Ed.), Test design contribution from psychology, education, and psychometrics. New York: Academic Press, in press.

Pintner, R. Intelligence and its measurement: A symposium. Journal of Educational Psychology, 1921, 12, 123-147; 195-216.

Pachella, R. G. The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), Human information-processing: Tutorials in performance and cognition. Hillsdale NJ: Erlbaum, 1974.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

Reitman, W. Cognition and thought. New York: Wiley, 1965.

Sternberg, R. J. Intelligence, information processing, and analogical reasoning. Hillsdale NJ: Erlbaum, 1977.

Sternberg, R. J. The nature of mental abilities. American Psychologist, 1979, 34, 214-230. (b)

Sternberg, R. J., & Weil, E. M. An aptitude strategy interaction in linear syllogistic reasoning. Journal of Educational Psychology, 1981, 73.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 1969, 4, 421-457.

Whitely, S. E. A general multicomponent latent trait model for aptitude processes . Paper presented at the annual meeting of Psychometric Society, Chapel Hill NC, May 1980. (a)

Whitely, S. E. Alternative strategy models for analogical reasoning. Paper presented at the annual meeting of the Psychometric Society, St. Louis MO, November 1980. (b)

Whitely, S. E. Latent trait models in the study of intelligence. Intelligence, 1980, 4, 97-132. (c)

Whitely, S. E. Multicomponent latent trait models for ability tests. Psychometrika, 1980, 45, 479-494. (d)

Whitely, S. E. Measuring aptitude processes with multicomponent latent trait models. Journal of Educational Measurement, 1981, 18, 67-84.

Whitely, S. E. The construct representation and nomothetic span of verbal ability. Paper presented at the annual meeting of the American Psychological Association, Washington DC, August 1982.

Whitely, S. E., & Barnes, G. M. The implications of processing event sequences for theories of analogical reasoning. Memory and Cognition, 1979, 7, 323-331.

Whitely, S. E., & Nieh, K. Program MULTICOMP. Unpublished manuscript, University of Kansas, Lawrence KS, 1981.

Whitely, S. E., & Schneider, L. M. Information structure on geometric analo-

gies: A test theory approach. Applied Psychological Measurement, 1981, 5, 383-397.

Woodrow, H., Intelligence and its measurement: A symposium. Journal of Educational Psychology, 1921, 12, 123-147; 195-216.

ACKNOWLEDGMENTS

The author would like to thank Lisa Schneider for preparing the figures and for her help in analyzing the data. This research was partially supported by National Institute of Education Grant No. NIE 6-7-1056 to Susan E. Whitely, Principal Investigator. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.